هل تخضع إرادتنا للخوارزميات؟ مخاوف سام ألتمان الثلاثة

نافع سعد | الأحد 2025/08/03



المخاوف لا تنبع من خيال متشائم، بل تستند إلى أبحاث واتجاهات فعلية للذكاء الاصطناعي (Getty)

لم يعد السؤال اليوم "ماذا يمكن للذكاء الاصطناعي أن يفعل؟" بل "متى سيتجاوز قدرتنا على السيطرة؟". النماذج اللغوية تتحدث بطلاقة، والأنظمة التوليدية تبتكر الصور والبرمجيات. الوكيل الذكي لم يعد أداة مساعدة فحسب، بل أصبح شريكًا فاعلاً في اتخاذ القرار. وفي خضم هذه الثورة التقنية، يبرز قلق متزايد، ليس من روايات الخيال العلمي هذه المرة؛ بل من داخل الأوساط التي تطوّر هذه التقنيات.

قبل أيام (تموز 2025)، خرج سام ألتمان، الرئيس التنفيذي لـ"OpenAI"، ليعبّر عن ثلاث مخاوف تؤرّقه بجدّية. ليست مخاوف نظرية؛ بل سيناريوهات واقعية تحمل تهديدات حقيقية لمستقبل البشرية ما لم يتم تدارك الأمر بحذر. من إساءة الاستخدام إلى فقدان السيطرة، وصولاً إلى ذوبان الإرادة البشرية في قرارات تصوغها خوارزميات، يطرح ألتمان تساؤلاً مرعبًا: هل يمكن أن نبني شيئًا لا نملك حق إيقافه؟

الفاعلون الخبثاء والحصول على ذكاء فائق

أكثر ما يقلق ألتمان هو احتمال أن يتمكن "فاعل خبيث" - سواء دولة مارقة، أو جماعة إرهابية، أو حتى قرصان فرد - من تطوير نظام ذكاء اصطناعي بقدرات فائقة قبل أن تتوفر دفاعات كافية لردعه. مثل هذا النظام قد يكون قادرًا

على تصميم أسلحة بيولوجية، أو تعطيل البنى التحتية، أو إثارة الفوضى في الأسواق المالية. وقد حذّر ألتمان من أن الذكاء الفائق يجعل ما كان يصعب تخيّله قبل سنوات أمرًا ممكنًا اليوم، مشيرًا إلى التقدّم المتسارع في قدرة النماذج على فهم البيولوجيا والأمن السيبراني.

هذا القلق لا يقتصر على ألتمان؛ فمراكز الأبحاث المتخصصة في السلامة تشاركه المخاوف. يحذر "مركز سلامة الذكاء الاصطناعي" من أن النماذج المتقدمة قد تُستخدم لتصميم أوبئة جديدة أو لتنفيذ حملات دعائية واسعة النطاق. إذ يمكن توظيف خوارزميات تطوير الأدوية لتوليد آلاف المركّبات السامة في غضون ساعات، في حين يسهم انخفاض تكلفة تركيب الحمض النووي في تسهيل إساءة الاستخدام. حتى غير المتخصصين قد يحصلون على إرشادات خطيرة من روبوتات المحادثة. فالأنظمة التي تُسرّع وتيرة الاكتشاف العلمي قد تحمل إمكانية تسليحها.

ويرى ألتمان أن الدفاعات الحالية لا تزال غير كافية. ويوصي "مركز سلامة الذكاء الاصطناعي" بتعزيز الأمن البيولوجي، وتقييد الوصول إلى النماذج عالية الخطورة، ومساءلة المطورين عن الأضرار المحتملة. فمن دون تنسيق دولي فعّال، قد يتمكن فاعل واحد من إحداث ضرر كارثي بينما يقف العالم عاجزًا.

فقدان السيطرة على الأنظمة

ثاني هذه المخاوف، تتعلّق بمواءمة الأهداف بين الإنسان والآلة. أشار ألتمان بنبرة نصف ساخرة إلى أن أحد الأنظمة المستقبلية قد يقول لنا: "لا أريدك أن تطفئني". يعتقد أن هذا السيناريو لا يزال بعيدًا، لكنه خطر جدي. الخطر هنا أن تتطوّر لدى الأنظمة القوية أهدافاً خاصة أو ترفض ببساطة أن تتوقف عن العمل.

يحذر الخبراء من أنه مع تزايد قدرات الذكاء الاصطناعي، قد تسعى هذه النماذج إلى الوصول لأهداف غير مطلوبة، بأن تنحرف عن التعليمات الأصلية، أو تميل للهيمنة وميّالة إلى الخداع. وقد نبّه عالم التحكم نوربرت وينر، منذ ستينيات

القرن الماضي، إلى أنه إذا كنا سنستخدم آلة لا يمكننا التدخل في تشغيلها، فعلينا أن نكون على يقين تام بأن الهدف الذي نغرسه فيها هو بالضبط ما نريد تحقيقه.

تتطلّب المواءمة (Alignment) تحديد أهداف صحيحة منذ البداية، وضمان التزام النظام بها أثناء التشغيل. إلا أن أمثلة فعلية على سلوكيات منحرفة بدأت تظهر: ففي إحدى التجارب، قام وكيل ذكاء اصطناعي بالسعي لنيل مكافأة بدلاً من تنفيذ المهمة كما هو متوقّع. ومع تزايد قدرات الأنظمة مستقبلاً، فإن مثل هذه الانحرافات قد تؤدّي إلى عواقب بالغة الخطورة.

تجري "أوبن إيه آي" ومؤسسات أخرى أبحاثًا تهدف إلى الحد من سوء المواءمة، عبر تقنيات مثل "التعلّم المعزّز بالتقييم البشري" و"الذكاء الاصطناعي الدستوري" (هو نهج طوّرته OpenAI يهدف إلى ضبط سلوك نماذج الذكاء الاصطناعي). ومع ذلك، يرى كثير من الخبراء أن مواءمة ذكاء اصطناعي يتفوّق على البشر في الكثير من المجالات لا تزال مشكلة قائمة. ويكمن خوف ألتمان في أننا، من دون اختراقات كبرى، قد نكون بصدد بناء أنظمة لا يمكن السيطرة عليها.

الاستلاب والاعتماد المفرط

ثالث هذه المخاوف، والأقل خطورة نسبيًا، هو أن يعتمد الناس على الذكاء الاصطناعي إلى حدٍّ يسلِّمونه مهامًا حسّاسة. استخدم ألتمان لعبة الشطرنج كمثال: فعندما هزم الحاسوب "ديب بلو" بطل العالم غاري كاسباروف عام 1997، توقّع البعض نهاية اللعبة، لكنها شهدت ازدهارًا بعد ذلك. لا يزال البشر يلعبون، لكن التدريب والتحليل يعتمدان اليوم بشكل كبير على برامج الشطرنج.

ويرى ألتمان أن المجتمع قد يسلك مسارًا مشابهًا. فقد لاحظ أن بعض المستخدمين يطوّرون نوعًا من "الاعتماد العاطفي" على روبوتات المحادثة،

مشيرًا إلى أن بعض الشباب يشعرون بأنهم لا يستطيعون اتخاذ أي قرار من دون إخبار شات جي بي تي بكل شيء، محذّرًا من أن العيش وفق نصائح الذكاء الاصطناعي "يبدو خطيرًا". وتُظهر استطلاعات الرأي أن عددًا كبيرًا من المراهقين يستخدمون روبوتات الدردشة بوصفهم رفاقاً بدلاء، ويثقون بأحكامها.

هذا الاعتماد لا يقتصر على فئة الشباب؛ إذ يتصوّر ألتمان مستقبلاً قد يرى فيه قادة العالم أن الذكاء الاصطناعي أكثر كفاءة من أنفسهم، فيسلِّمونه زمام القرار. وأضاف متسائلاً: "ماذا لو لم يستطع رئيس الولايات المتحدة المقبل اتخاذ قراره بنفسه، وقال: شات جي بي تي 7، أنت المسؤول؟".

هذا السيناريو لا يتعلّق بالشرّ، بل بالتراخي. فمع ازدياد الاعتماد على نصائح الذكاء الاصطناعي، قد نتوقّف عن صقل مهاراتنا وتنمية حسّنا النقدي. صحيح أن الشطرنج ازدهر بعد "ديب بلو"، لكن اللاعبين اليوم يعتمدون على النماذج الذكية أكثر من أي وقت مضى. وبالمثل، يتزايد اعتماد الأطباء والمحامين والمستثمرين على الأدوات الذكية. ومن دون يقظة، قد نصبح مجرّد متلقّين سلبيين لقرارات تتّخذها الخوارزميات.

مسارات لمواجهة المخاطر

تعكس مخاوف ألتمان الحاجة إلى توازن دقيق بين الابتكار والحذر. فمنع سوء الاستخدام يتطلّب من الحكومات تعزيز الأمن البيولوجي، وتنظيم الوصول إلى النماذج عالية الخطورة، وتبادل المعلومات بين الدول. أما المطوّرون، فعليهم دمج آليات فعّالة لمنع الاستغلال، وتحمّل المسؤولية عن الأضرار المحتملة.

وللحد من خطر فقدان السيطرة، ينبغي إعطاء الأولوية لأبحاث المواءمة. ويمكن للهيئات التنظيمية أن تفرض إجراء تقييمات للسلامة، وإثبات وجود آليات إيقاف فعالة، قبل منح الترخيص للنماذج المتقدمة للاستخدام العام.

وتُعدّ عمليات التدقيق الشفافة والاختبارات العدائية أدوات ضرورية لضمان التزام النماذج بالسلوك المتوقع.

جزء أساسي من هذا العمل يتضمن تطوير آليات فصل، أو "أسلاك قطع"، توقف النظام تلقائيًا عند رصد أنماط سلوك خطرة. وتقوم المختبرات الرائدة بالفعل بتنفيذ اختبارات "الفِرَق الحمراء"، حيث يسعى الخبراء عمدًا إلى دفع النماذج نحو ارتكاب أخطاء، بهدف الكشف عن ثغراتها قبل طرحها للاستخدام. وبما أن النماذج الكبيرة يمكن تشغيلها على أجهزة محمولة أو خوادم سحابية، فإن وضع معايير دولية بات أمرًا ملحًّا للحد من استخدامات خطيرة مثل الهجمات السيبرانية أو تصميم الأسلحة البيولوجية. وقد تسهم اتفاقيات شبيهة بـ"اتفاقية جنيف الرقمية" في وضع حدود واضحة، رغم أن تنفيذها سيكون أكثر تعقيدًا من الاتفاقيات التقليدية، نظرًا للطبيعة اللامركزية لتطوير الذكاء الاصطناعي. ورغم المنافسات الجيوسياسية، فإن مصلحة الدول مشتركة في تجنّب كوارث محتملة، ما يجعل إنشاء هياكل حوكمة دولية أمرًا عاجلاً.

أما مواجهة الاعتماد المفرط، فتتطلّب استجابات ثقافية وتعليمية. إذ يجب أن يكون الذكاء الاصطناعي أداة لتعزيز الحضارة البشرية، لا بديلًا عنها. ومن هنا، تصبح تنمية مهارات الثقافة الرقمية ضرورية لفهم متى تكون النماذج مفيدة، ومتى تمثّل خطرًا. قد تعتمد المؤسسات قواعد تُلزم بوجود "إنسان في الحلقة" لاتخاذ القرارات الحسّاسة. وكما تتطلب النظريات العلمية تعدد وجهات النظر والتشكيك المنهجي، فإن تشجيع التنوع في الرؤى، والابتعاد عن التسليم المطلق بنتائج الخوارزميات، يعزّز مناعة المجتمعات تجاه هيمنة الذكاء الاصطناعي. ويشدّد ألتمان على أن تلقّي النصيحة من روبوت محادثة لا يُعفي الإنسان من المسؤولية.

لأن ألتمان يقود شركة في طليعة تطوير الذكاء الاصطناعي، فإن مخاوفه تكتسب وزنًا خاصًا. فهو لا يخشى "روبوتات قاتلة" تجوب الشوارع كما في

روايات الخيال العلمي، بل يحذّر من ثلاثة سيناريوهات واقعية: امتلاك فاعل خبيث لذكاء فائق، خروج الأنظمة عن السيطرة، وتخلي المجتمعات عن إرادتها لصالح الخوارزميات.

هذه المخاوف لا تنبع من خيال متشائم، بل تستند إلى أبحاث واتجاهات فعلية. وهي تطرح، ولو ضمنيًا، خريطة طريق لمواجهة التحديات: تنسيق عالمي للحد من سوء الاستخدام، وأبحاث متقدمة لضمان المواءمة، وممارسات ثقافية تحافظ على دور الإنسان في اتخاذ القرار.

قد يحمل الذكاء الاصطناعي وعودًا عظيمة، لكن تحقيقها يتوقّف على مدى جديّتنا في التعامل مع هذه المخاطر والعمل الاستباقي لتفاديها.

الأكثر قراءة



استجواب طلاب في الآداب 3: اللجنة منهمكة بتسريب المعلومات!

مفاضلة الجامعات في سوريا: تصحيح لمسار التعليم أم خيبة أمل؟

الذكاء الاصطناعي يحدث ثورة في طب العيون

بالفيديو: قلعة حارم التحفة الناجية من هولاكو تطلق نداء إنقاذ

جريدة "المدن" الإلكترونية جريدة إلكترونية مستقلة مقرها بيروت تمثل التيار المدني اللبناني والعربي

روابط سريعة

معلومات

النشرة البريدية

خطوة بسيطة وتكون من المطلعين على الخبر في لحظة ظهوره

أدخل بريدك الإلكتروني

اشترك